Signal Processing 21 (1990) 17-24
Elsevier

A MAXIMUM ENTROPY APPROACH TO INTERPOLATION

S. Tanveer FATHIMA* and B. YEGNANARAYANA

Department of Computer Science and Engineering, Indian Institute of Technology, Madras, India 600036

Received 27 August 1987
Revised 20 February 1989 and 24 January 1990

Abstract. The maximum entropy approach has been applied to several problems including spectrum estimation, image reconstruction, etc. In this paper we use this approach to address the interpolation problem. Specifically, we address the problem of recovery of the missing samples of a signal from a few of its randomly distributed samples. We also discuss the appropriateness of the maximum entropy method for different distributions of the known samples. Finally, we extend this method to handle additional information about the signal, if available.

Zusammenfassung. Der Maximum-Entropie-Ansatz ist auf verschiedenartige Probleme angewandt worden wie etwa die Spektralschätzung, die Bildrekonstruktion, usw. Im folgenden verwenden wir diesen Ansatz, um das Interpolationsproblem anzugehen. Insbesondere wenden wir uns der Aufgabe zu, die fehlenden Abtastwerte eines Signals aus wenigen, in zufälligen Punkten gegebenen Werten wiederzufinden. Wir diskutieren auch die Eignung des Maximum-Entropie-Ansatzes bei unterschiedlichen Verteilungen der bekannten Abtastwerte. Schließlich dehnen wir dieses Verfahren aus auf die Ausnützung zusätzlicher, unter Umständen verfügbarer Informationen über das Signal.

Résumé. Le principe du maximum d'entropie a été appliqué à divers problèmes parmi lesquels l'estimation spectrale, la reconstruction d'image, etc. Dans cet article, cette approche est utilisée pour l'interpolation. On étudie plus précisément le problème de la récupération de points manquants d'un signal à partir de quelques échantillons distribués aléatoirement. L'opportunité de la méthode du maximum d'entropie pour diverses distributions des échantillons connus est également discutée. Finalement, cette approche est étendue pour exploiter d'éventuelles informations supplémentaires sur le signal.

Keywords. Entropy, interpolation, reconstruction

1. Introduction

The problem of recovery of a signal from partial information arises in several situations. In noisy environment as in radar, or during measurements taken using faulty instruments, the signal may be so corrupted that some of its values may have to be discarded. The data here consists of a few received signal values that are reliable. Usually, in such situations, some a priori information is also available. For example, we have the bandlimited signal interpolation problem [11] (with bandwidth known), the reconstruction from phase problem [12] (duration of the signal known), etc. In this paper, however, we are interested in the recovery of a signal from only a few of its randomly distributed samples. No a priori information is assumed to be available. In other words, we address the traditional interpolation problem. Interpolation has been attempted using several different approaches. These include polynomial interpolation schemes (such as Lagrangian, spline, etc.), nonuniform sampling methods [13] and model-based approaches [5]. All these methods make some assumption about either the number of samples known (see [5]), or the distribution of

* Current Address: Artificial Intelligence Laboratory, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

these samples (such as a Poisson distribution as in [13]). This phenomenon is also observed in the maximum entropy spectrum estimation methods [10]. Since these methods do not explicitly incorporate the frequency information that may be present in the given data (i.e., in the known samples) during interpolation, they often cause a low frequency bias. Also, since the recovery of the missing samples is not possible with the given information (it is an underdetermined problem), these methods try to give only an estimate of the solution.

In this paper we consider yet another method of obtaining an estimate of the signal based on the given data. We look for an estimate that is not only consistent with all the available data but is also maximally noncommittal about the missing information. In other words, we look for a maximum entropy approach to interpolation. The concept of maximising the entropy to get a solution is not new. The rationale behind such a problem solving methodology was given by Jaynes [6]. It has been applied to several areas including spectrum estimation [3], image reconstruction [4], etc.

The general strategy behind applying the maximum entropy method has been as follows. The procedure begins with the building of a statistical model of the process at hand. A measure of entropy is then chosen. This measure indicates the randomness or uncertainty in the statistical environment. All a priori knowledge about the problem is then expressed as constraints. The entropy expression is then combined with the constraints using the Lagrangian multiplier method. This combined function, in effect, represents our total awareness of the problem, since it expresses the known information as well as our ignorance about the unknown information. This is then maximised with respect to the unknown variables of the system to obtain a solution for these variables.

Thus any new application of the maximum entropy (ME) method entails doing the above operations for the given problem domain. In this paper we show how the interpolation problem can be addressed using the above formulation. The paper is organized as follows. In Section 2 we obtain an ME estimate for the signal. In Section 3 we present an algorithm for obtaining such an estimate. We discuss implementation aspects in Section 4 and present some experimental results. In Section 5 we discuss the role of the known signal samples in generating the ME estimate. Finally, in Section 6 we extend the method to include other a priori information that may be available.

2. Interpolation using ME method

In interpolation, the missing samples that are in between the known samples of a deterministic signal are to be recovered. Therefore the concept of a most random estimate for such samples (using the ME principle) seems counterintuitive. However, the following stochastic model for this deterministic process justifies the ME approach:

For any given distribution of the known samples, several signals are possible that have these known samples, but have completely different values at the locations of the unknown samples. All these signals then constitute the sample functions of a random process. In this process, a probability p_{ij} is associated with a signal sample at instant i, to denote the likelihood of it taking a value x_j, where x_j is any real number. The desired signal is thus one of a number of possible signals in this random process. Since it is difficult to choose it exactly, one way is to settle for an estimate of the desired signal. This estimate could be such that it agrees with all the known sample values, and at the same time remains maximally noncommittal with regard to the unknown values. To obtain such an estimate, the uncertainty about the unknown values should first be expressed in a measure. One such uncertainty measure is the Shannon's measure: the entropy. It is given as

$$H = -\int_{-\infty}^{\infty} p(u) \log p(u) \, du, \tag{1}$$

where $p(u)$ is the probability distribution of the signal amplitude. Since $p(u)$ is not available, we circumvent this difficulty by relying on a difference in entropy rather than the absolute entropy. This approach is similar to the one used for spectrum estimation [1]. Specifically, we use the linear filter method of [1]. For this, we assume that the signal spectrum decays after some frequency σ (this is the usual assumption made while selecting the sampling rate). Then we can regard the signal spectrum to be bandlimited to σ. That is

$$S(\omega)=0 \quad \text{for } |\omega|>\sigma. \tag{2}$$

Note that we are not considering the case of bandlimitedness as a priori information here. In that case the exact location and extent of the band would be known. This case will be considered later in Section 6. Also, note here that by signal spectrum we mean the Fourier transform of the signal and not its power spectrum.

Using the above information, and the linear filter method, the output entropy rate H_s can be expressed in terms of the input entropy rate H_g as (for details see [9])

$$H_s = H_g + \frac{1}{2\sigma}\int_{-\sigma}^{\sigma} \log|S(\omega)|\,d\omega. \tag{3}$$

Even though we now have the entropy rate, it does not matter since maximising the entropy rate is equivalent to maximising the entropy itself (for a fixed sample size). The difference between the input and output entropy rates is given by

$$\Delta H = \frac{1}{2\sigma}\int_{-\sigma}^{\sigma} \log|S(\omega)|\,d\omega. \tag{4}$$

For the output $s(t)$ to have as great an entropy as possible, ΔH, the so-called entropy gain of the filter, should be made as large as possible. Therefore we choose ΔH as our entropy measure I.

Expressing the entropy in terms of the DFT samples of the signal spectrum, we have

$$I = \Delta H = K_1 \sum_{k=-M}^{M} \log|S(k)|, \tag{5}$$

where $K_1 = \pi/N\sigma$, $M = \lfloor N\sigma/2\pi \rfloor$, N is the number of DFT points and $S(k)$ are the DFT samples of $S(\omega)$.

At this point we notice that the entropy expression involves the signal spectrum rather than the signal itself. But this again does not matter since recovering the Fourier spectrum of the signal is equivalent to recovering the signal itself.

The constraints in the interpolation process are provided by the known signal samples. They are related to the signal's spectrum as

$$Q_i : K_2 \sum_{k=-M}^{M} S(k)\exp\left(\frac{j2\pi t_i k}{N}\right) - s(t_i) = 0, \tag{6}$$

where $1 \leqslant i \leqslant L$, $K_2 = 1/N$ and $s(t_i)$ are the known signal samples at instants t_i. Here L samples are assumed to be known.

We now use the Lagrangian multiplier method to solve for $S(k)$. I in (5) can be expressed as a function of $2M+1$ variables $S(k)$ as

$$I = \Phi(S(-M), \ldots, S(M)). \tag{7}$$

Note that both I and Q_i are real functions of the complex variables $S(k)$. Now form a function Ψ given by

$$\Psi(S(-M), \ldots, S(M)) = I + \sum_{i=1}^{L} \lambda_i Q_i, \tag{8}$$

where λ_i are the Lagrangian multipliers. Maximising Ψ with respect to the variables $S(-M), \ldots, S(M)$ yields

$$S(k) = -\frac{K_3}{\sum_{i=1}^{L} \lambda_i \exp\left(\frac{j2\pi t_i k}{N}\right)}, \tag{9}$$

where $-M \leqslant k \leqslant M$ and $K_3 = K_1/K_2$.

Using (6) and (9), we get

$$s(t_p) = -K_1 \sum_{k=-M}^{M} \frac{\exp\left(\frac{j2\pi t_p k}{N}\right)}{\sum_{i=1}^{L} \lambda_i \exp\left(\frac{j2\pi t_i k}{N}\right)}, \tag{10}$$

where $1 \leqslant p \leqslant L$.

Equation (10) is a set of simultaneous nonlinear equations in λ_i. The λ_is solved using (10) can then be substituted in (9) to get $S(k)$. Finally, by performing the inverse DFT, we can recover all the samples of the ME estimate as

$$s(n) = -K_1 \sum_{k=-M}^{M} \frac{\exp\left(\frac{j2\pi nk}{N}\right)}{\sum_{i=1}^{L} \lambda_i \exp\left(\frac{j2\pi t_i k}{N}\right)}, \quad (11)$$

where $0 \leq n \leq N-1$.

3. Algorithm for obtaining the MEM signal estimate

The algorithm to obtain the maximum entropy estimate of the signal when some of its randomly distributed samples are known, is given below. It essentially uses Broyden's method for solving nonlinear simultaneous equations [2] extending it to handle complex variables, since the langrangian multipliers will in general be complex.

1. Construct a signal $y_0(n)$ as follows:

$$y_0(n) = \begin{cases} s(n), & n \in I_T \text{ and } 0 \leq n \leq N-1, \\ 0, & n \notin I_T \text{ and } 0 \leq n \leq N-1. \end{cases} \quad (12)$$

2. Compute the Fourier transform of $y_0(n)$, as $Y_0(k)$.
3. Take L samples of $Y_0(k)$.
4. Solve for the initial estimates of λ_i^0 using the following set of linear equations obtained from the L samples of $Y_0(k)$ chosen in Step 3 as

$$\sum_{i=1}^{L} \lambda_i^0 \exp\left(\frac{j2\pi t_i k}{N}\right) = -\frac{K_1}{Y_0(k)}. \quad (13)$$

   The above equations can be solved using Gaussian elimination.
5. Compute A_0 the Jacobian $[\partial f_i / \partial \lambda_j^0]$, $1 \leq i, j \leq L$, where

$$f_i = s(t_i) + K_1 \sum_{k=-M}^{M} \frac{\exp\left(\frac{j2\pi t_i k}{N}\right)}{\sum_{l=1}^{L} \lambda_l^0 \exp\left(\frac{j2\pi t_l k}{N}\right)}. \quad (14)$$

6. Compute $K_0 = A_0^{-1}$.
7. Obtain the vector $F(\Lambda^0)$ (where $\Lambda = [\lambda_i]$) as

$$F(\Lambda^0) = s(t_p) + K_1 \sum_{k=-M}^{M} \frac{\exp\left(\frac{j2\pi t_p k}{N}\right)}{\sum_{i=1}^{L} \lambda_i^0 \exp\left(\frac{j2\pi t_i k}{N}\right)} \quad (15)$$

   for $1 \leq p \leq L$.
8. Set $m = 1$.
9. Update the Lagrangian multiplier vector Λ^m as follows:

$$\Lambda^{m+1} = \Lambda^m - K_m * F(\Lambda^m). \quad (16)$$

10. Compute the difference vector $\Delta\Lambda$ as

$$\Delta\Lambda = \Lambda^{m+1} - \Lambda^m. \quad (17)$$

11. Compute the vector $F(\Lambda^{m+1})$ as in Step 7.
12. Compute the difference vector ΔF as

$$\Delta F = F(\Lambda^{m+1}) - F(\Lambda^m). \quad (18)$$

13. Update the inverse of the Jacobian K_m as

$$K_{m+1} = K_m - \frac{[K_m * \Delta F - \Delta\Lambda] * (\Delta\Lambda)^{\mathrm{T}} * K_m}{(\Delta\Lambda)^{\mathrm{T}} * K_m * \Delta F}. \quad (19)$$

14. Repeat Steps 9 through 13 until convergence is reached (i.e., $F(\Lambda^m) \approx 0$). In other words, the absolute value of each element of the vector $F(\Lambda^m)$ should be 0.

From the above description of the algorithm, it is evident that it takes $O(L^3)$ operations for Step 6, while for Steps 9 through 13, it takes only $O(L^2)$ operations per iteration.

We now discuss some issues that arise in solving for the ME estimate using (10). The algorithm given above uses an iterative method to solve (10). As we know, such methods require good initial estimates (here for the λ_i values) to converge to a solution. In general, for large number of constraints (and hence a large number of λ_is), it becomes difficult to choose good initial estimates. Several ad hoc methods are used for this purpose (see for example [7], where the problem of choosing the initial estimates for λ_i and the autocorrelation values is discussed for the 2D power spectrum estimation problem). We have chosen to find the initial estimate from the given data, and any a priori information that may be available. Specifically, we use the Fourier spectrum of the data to obtain the initial estimates. This is justified due to the following observation. It has been found that even with as many as 90% of the signal samples missing, the data spectrum, in most cases, does indicate the true spectrum, especially the peaks in the signal's spectrum [11]. Therefore, we make use of L samples of the data spectrum $S(k)$ from $k=-\frac{1}{2}L+1$ to $\frac{1}{2}L$ to solve for the L λ_i values using (9). However, other choices of these L spectrum samples are also possible. For example, L samples can be chosen around the peaks in the data spectrum.

Using the data spectrum has also another advantage. It allows any frequency information inherent in the given data to be incorporated in the interpolation procedure. This together with the fact that this procedure gives a maximally noncommittal estimate is the reason why the resulting interpolated signal does not suffer from low frequency bias.

4. Results

We now illustrate the ME method by an example. Consider the low pass signal shown in Fig. 1(a). The bandwidth of this signal is given in Table 1. Knowing σ and N, the number of DFT points, the constants M, K_1, K_2, K_3 can be evalu-

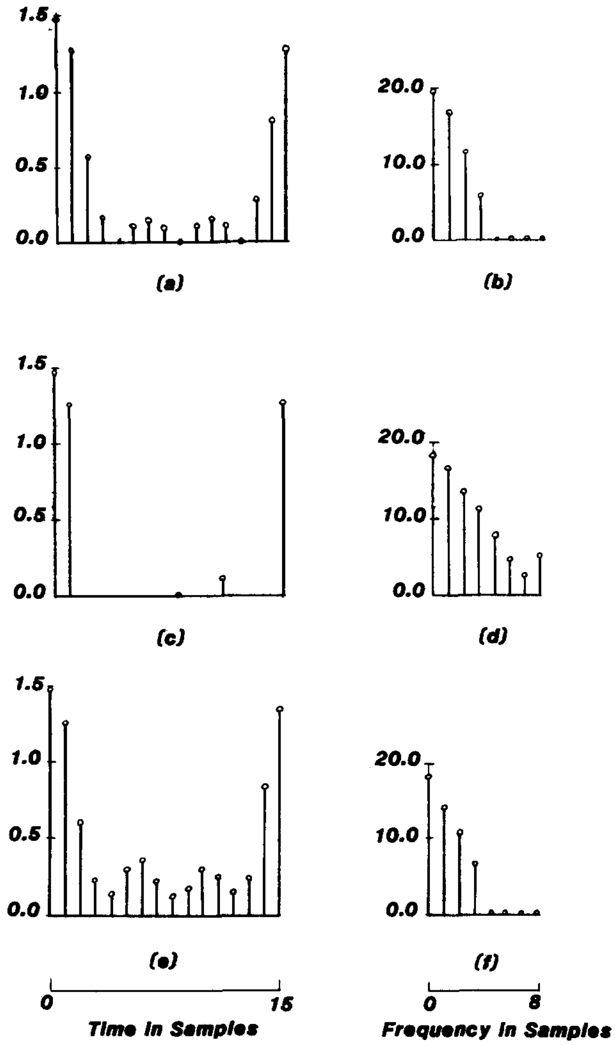

Fig. 1. Illustration of signal reconstruction from partial data using the maximum entropy method—Randomly distributed samples case. (a) Original signal. (b) Spectrum of the original signal. (c) Given partial data. (d) Spectrum of the partial data. (e) Reconstructed signal after 8 iterations. (f) Spectrum of the reconstructed signal.

Table 1

Parameters used for the illustration

band-width σ	DFT points N	band extent M	signal length	known samples L	K_1	K_2	K_3
0.375 π	16	±3	16	5	0.166	0.0625	2.67

ated using their definitions given in (5), (6) and (9), respectively. These parameters are also indicated in Table 1. The signal duration is assumed to be 16 samples. Five signal samples are available as the known data. Figure 1(c) shows a possible known sample distribution. The reconstructed signal using the ME method, obtained after 8 iterations is shown in Fig. 1(e). Its spectrum is shown in Fig. 1(f).

Figure 2(a) shows another data distribution of the same signal of Fig. 1(a). This is an instance of the extrapolation problem. The reconstructed signal obtained after 10 iterations is shown in Fig. 2(c).

The above two examples were for the case where the signal was even, that is, the Fourier transform was real. Figure 3 demonstrates the effectiveness of the method for the case where the Fourier transform is complex. The parameters used are again

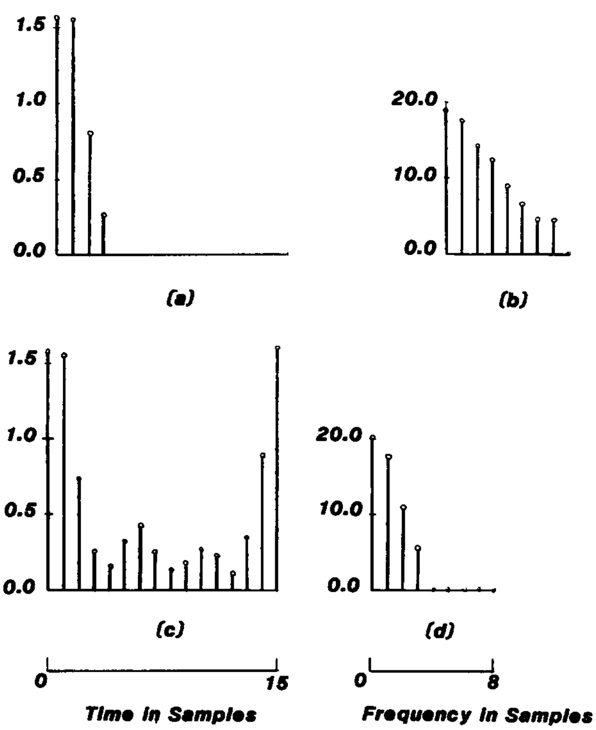

Fig. 2. Illustration of signal reconstruction from partial data—An extrapolation case. (a) Given partial data for the original signal of Fig. 1(a). (b) Spectrum of the data. (c) Reconstructed signal using ME method after 10 iterations. (d) Spectrum of the reconstructed signal.

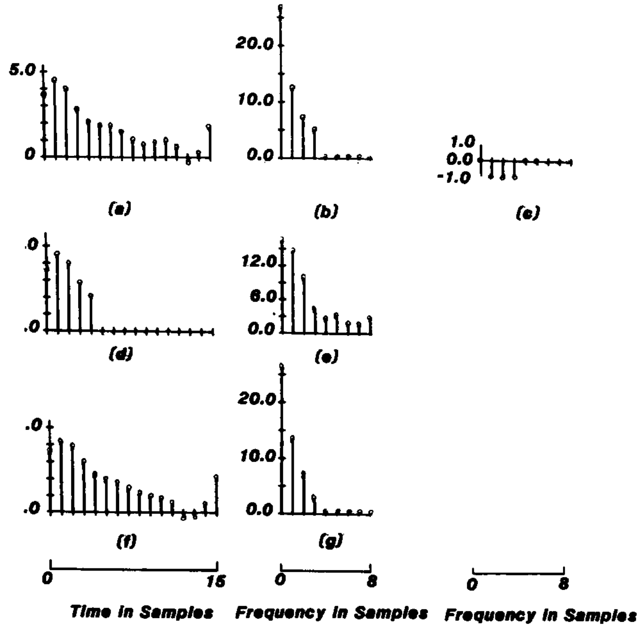

Fig. 3. Illustration of signal reconstruction from partial data—A nonlinear phase spectrum case. (a) Original signal. (b) Magnitude spectrum of the signal. (c) Phase spectrum of the signal. (d) Given partial data for the signal of Fig. 3(a). (e) Magnitude spectrum of the given data. (f) Reconstructed signal after 5 iterations. (g) Magnitude spectrum of the reconstructed signal.

as indicated in Table 1. Here again, the duration of the signal is 16 samples. Notice in Fig. 3(c) that the phase spectrum of the signal is nonlinear.

After trying several different signals we found that, in a number of cases, the ME method converges to an estimate that closely resembles the original signal (i.e., with a low mean square error $\approx 2 * 10^{-4}$). These results, however, must be interpreted carefully, as we show in the next section.

5. Discussion

We now discuss the implications of using the ME approach for interpolation. From the formulation given in Section 2, it is clear that there is no restriction on the distribution of the known samples. Thus extrapolation, uniform distribution, as well as other distributions can be handled in a uniform manner. This is borne out in Figs. 1(e) and 2(c). An inspection of these figures also reveals that they resemble the original signal in Fig. 1(a). However, this is not always the case, because the ME approach does not aspire to reconstruct the signal exactly. It only gives an estimate that is guaranteed to be the most noncommittal one, given the known information. Thus any comparison with the original signal is not really appropriate. This also raises the question of when this approach is then suitable. Clearly, when a large number of samples are known, it is reasonable to assume that the unknown samples can be predicted using them. For example, it is known that the samples of an over-sampled bandlimited signal are dependent on each other [8]. That is, the known samples can provide information about the missing samples. Then any approach such as the polynomial interpolation, or relaxation scheme, which makes use of this dependency will work just as well, or even better (see [11] for an algorithm that employs this dependency for signal recovery). On the other hand, if there are only a few samples known and they are arbitrarily distributed (possible far apart), and if the signal is not bandlimited, the traditional interpolation schemes using the dependency principle do not work well. In such cases, the ME approach, which allows an element of randomness in the unknown samples, can be recommended. This has also been verified in our experiments.

6. Extension of ME estimate

So far we have assumed that the only available information about the signal is a set of randomly distributed signal samples. If now some additional information is known, it can also be incorporated as long as it can be expressed as constraints. To see this, consider the case when a few spectrum values are known in addition to the known signal samples. These spectrum values can be incorporated as part of the constraints in (6), so that

$$Q_i : K_2 \sum_{k \notin I_F, |k| \le M} S(k) \exp\left(\frac{j2\pi t_i k}{N}\right) + K_2 \sum_{k \in I_F} S(k) \exp\left(\frac{j2\pi t_i k}{N}\right) - s(t_i) = 0, \quad (20)$$

where $1 \le i \le L$ and I_F is the set of frequency points at which the signal spectrum is known. Equation (10) is similarly modified as

$$s(t_p) = K_2 \sum_{k \in I_F} S(k) \exp\left(\frac{j2\pi t_p k}{N}\right) - K_1 \sum_{k \notin I_F, |k| \le M} \frac{\exp\left(\frac{j2\pi t_p k}{N}\right)}{\sum_{i=1}^{L} \lambda_i \exp\left(\frac{j2\pi t_i k}{N}\right)}, \quad (21)$$

where $1 \le p \le L$.

From the above equation we notice that only $L\,\lambda_i$ values have to be solved. Thus even though the constraints have increased, the solution does not become more complex since the extra constraint is integrated as part of Q_i itself. The ME estimate will now be different since it will not only agree with the known signal samples but also with the known spectrum samples. Also, since more is known about the signal than before, the total uncertainty and hence the entropy is reduced. But with respect to the currently available information, the entropy is still the maximum.

We can similarly incorporate other information if it can also be expressed as a constraint. Thus if it is known that the spectrum of the signal is bandlimited to a given frequency band, and that it is low pass in nature, it can be easily incorporated into the above analysis by changing the value of the parameter M (M will now be $\lfloor N\sigma_n/2\pi \rfloor$, where σ_n is the actual extent of the band). Everything else is the same except that there are fewer spectrum values to be solved using (9). If the spectrum is bandlimited but of the bandpass type, then the limits of the integral in (3) have to be changed appropriately.

Thus we see that when some a priori knowledge is available, it can be incorporated into the estimate as long as it can be analytically expressed as a constraint. It is also interesting to see how such previous knowledge modifies the convergence of the algorithm for obtaining the ME estimate given in Section 3. If the extra information can be integrated well with the already existing constraints, then the cost of obtaining a solution does not go up. In that case, the solution is obtained faster. This we have seen for the specific case when the a priori information is some samples of the signal's Fourier transform. There the solution could be obtained faster as there were fewer variables to solve for. However, if the constraints are such that they result in complex computations for the same number of variables to solve for, then the updation of A in (16) takes a long time because of the longer time taken to update the inverse of the Jacobian in (19).

7. Conclusions

In this paper we have addressed the interpolation problem using the maximum entropy approach. We have given an ME estimate for a signal when a few of its samples are known. We have also discussed the appropriateness of the ME estimate for different data distributions and, specifically, its suitability for sparse data situations. Finally, we have extended the ME approach to handle more information without increasing the complexity of the solution process.

References

[1] J.G. Ables, "Maximum entropy spectral analysis", *Astronom. and Astrophys. Suppl. Series*, Vol. 15, 1974, pp. 383–393.

[2] C.G. Broyden, "A class of methods for solving nonlinear simultaneous equations", *Math. Comp.*, Vol. 91, 1965, pp. 577–593.

[3] J.P. Burg, "Maximum entropy spectral analysis", *Proc. 37th Meet. Soc. Exploration Geophysicists*, 1967.

[4] B.R. Frieden, "Restoring with maximum likelihood and maximum entropy", *J. Opt. Soc. Amer.*, Vol. 62, No. 4, April 1972, pp. 511–518.

[5] A.J.E.M. Janssen, R.N.J. Veldhuis and L.B. Vries, "Adaptive interpolation of discrete-time signals that can be modelled as autoregressive processes", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-34, April 1986, pp. 317–330.

[6] E.T. Jaynes, "On the rationale of maximum entropy methods", *Proc. IEEE*, Vol. 70, 1982, pp. 939-952.

[7] J.S. Lim and N. Malik, "A new algorithm for two-dimensional maximum entropy power spectrum estimation", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-29, No. 3, June 1981, pp. 401-413.

[8] R.J. Marks II, "Restoring lost samples from an oversampled bandlimited signal", *IEEE Trans. Acoust. Speech. Signal Process.*, Vol. ASSP-31, June 1983, pp. 752–755.

[9] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd Ed., McGraw-Hill, New York, 1984.

[10] N. Rozario and A. Papoulis, "Spectral estimation from nonconsecutive data", *IEEE Trans. Inform. Theory*, Vol. IT-33, No. 6, November 1987, pp. 889-894.

[11] B. Yegnanarayana and S.T. Fathima, "An algorithm for bandlimited signal interpolation", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Tokyo, April 1986.

[12] B. Yegnanarayana, S.T. Fathima and H. Murthy, "Reconstruction from Fourier transform phase with applications in speech analysis", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Texas, April 1987.

[13] J.L. Yen, "On nonuniform sampling of bandwidth-limited signals", *IRE Trans. Circuit Theory*, Vol. CT-3, December 1956, pp. 251–257.